Structural Thinking and Epistemic Injustice

Nadya Vasilyeva & Saray Ayala-López

A chapter in B.R. Sherman & S. Goguen (Eds.), *Overcoming epistemic injustice: Social and psychological perspectives*.

0. Intro

In response to the question "How can we fight (at least some forms of) epistemic injustice?" we propose what can be classified as a structural-psychological approach, one that targets how people think about structural factors. We review recent empirical results suggesting that a structural-psychological intervention is a promising way to address discriminatory behavior in general. Our proposal is driven by the growing evidence that the discriminatory behavior behind practices of epistemic injustice is strongly associated with people's psychological tendency to form what the psychological literature calls internalist representations of social (and natural) categories. To situate our proposal in the existent literature on both epistemic injustice and the general debate about interventions on and explanations of social injustice, we start by briefly reviewing two main types of interventions, and locating them in the general debate between individualistic and structural approaches to social justice. Next we review empirical literature on the cognitive processes that contribute to discriminatory behavior, focusing on instances of internalist thinking, such as psychological essentialism and inherence heuristic. Next, we introduce our proposal of promoting structural thinking as an intervention against discriminatory behavior in

general, based on counteracting people's tendency to represent social categories in internalist ways. Finally, we review the benefits structural thinking might bring to the fight against epistemic injustice.

1. Individualistic and structural approaches to counteracting epistemic injustice

Epistemic injustice is about an unjust distribution and management of epistemic resources. The unjust (not merely unfortunate) distribution and management of resources can be divided into three subcategories. Testimonial injustice occurs when someone does not recognize the interlocutors' knowledge, usually via faulty attributions of credibility; hermeneutical injustice has to do with hermeneutical lacunas, gaps in collective epistemic resources that make it hard for people in different epistemic communities to conceptualize, understand and/or communicate certain experiences (Fricker 2007); finally, contributory injustice takes place when knowers are wronged because interlocutors, due to a willful hermeneutical ignorance (Pohlhaus 2012), utilize prejudiced hermeneutical resources (Dotson 2012). Various practices of epistemic injustice often work together, forming wider patterns of unjust social dynamics.

Interventions to prevent or alleviate epistemic injustice, and in particular the specific breed of testimonial injustice, often target people's implicit biases and prejudiced beliefs that seem to drive discriminatory behaviors. Fricker's (2007) proposal on the virtue of testimonial justice is a good example of this approach. Roughly, it works as follows: when you adopt a critical, reflective attitude and become aware of the possibility that your epistemic judgements are being affected by an identity-prejudice

against a speaker, you need to neutralize its impact. Doing so implies the need to "shift intellectual gear" from the spontaneous heuristics people use when evaluating speakers (which are infused with prejudices) to an "active critical reflection" that would help people figure out how the prejudices might have distorted their judgement (91). This intervention paradigm has received criticisms, often pointing out that enforcing the virtue of testimonial justice is not feasible and/or counterproductive to the overall goal of eliminating testimonial injustice (see e.g. Sherman 2015).

One straightforward alternative to Fricker-style proposals is to shift attention from individual's minds and focus on fixing the environment shaping people's behavior. That is, instead of trying to upgrade our flawed minds, full of problematic associations between social groups and traits, we could turn to the correlations that exist in our corrupted society and are picked up by our minds as we form representations useful for navigating the world we inhabit. Let's unpack how the two alternatives relate to each other.

Fricker-style proposals locate the wrong within our minds, which make up problematic associations between social groups and traits, later used as shortcuts to reason about group members. The objective is to resist the wrong mental representations and get the social reality right. But as Sally Haslanger reminds us, many morally problematic attitudes do get the social reality right: "Women actually are more submissive than men; we are better caregivers than men; we are better at multi-tasking too" (Haslanger, 2017, 3), and African Americans are five to eight times more likely to be imprisoned than European Americans (Robertson, 2004). The key is not to deny this,

but to ask *why* the social reality takes this shape. As part of an answer to this question, Haslanger continues: "This is not to say that this is true "by women's nature" but because of the social history of gender" (Ibid 3); a similar comment can be made about racial inequalities in incarceration rates, appealing to the social history of race. That is, when we inquire into many of the correlations present in our social reality, we see patterns of inequalities and unjust social (structural) dynamics that constitute and cause them. This opens up a potential locus of intervention: the social dynamics that creates and maintains those correlations that our minds readily pick up on. That is, instead of intervening on the shortcuts our minds make up, let's instead fix the environment that provides inputs to our mental machinery. Once the problematic associations disappear from the environment, relying on mental shortcuts to reason about it would not lead to discrimination and injustice anymore. In line with this approach, psychologist Gerd Gigerenzer, an advocate of the benefits of cognitive heuristics, suggests that "[T]o improve moral behavior towards a given end, changing environments can be a more successful policy than trying to change beliefs or inner virtues" (Gigerenzer 2010, 530). Heuristics are not necessarily bad, and they are certainly an important cognitive tool. These shortcuts allow us to navigate the world fast and under conditions of uncertainty. This is one way our mind copes with a complex world. The environmental intervention approach avoids the challenge of fighting basic characteristics of human psychology. Instead, it proposes to engineer the social system in a way that helps us align our judgements and actions with our explicit beliefs and available evidence.

The two interventions reviewed above could be seen as reflecting the divide that we find in the philosophical debate about social justice more generally. There, it is common to find an opposition between individualistic and structural approaches. While Fricker-style interventions can be considered individualistic, the alternative environment-engineering intervention could be characterized as structural.[1] Much can be said about whether or not individualistic and structural approaches are really opposite. Recent developments acknowledge that individualistic and structural approaches are not necessarily in competition. For example, Robin Zheng (2018) merges the two approaches proposing to understand implicit biases to be a particular type of social structure, rather than ordinary individual attitudes. In turn, Lacey Davidson and Dan Kelly (ms) see norms, and our capacity to detect them and follow them, as the connection between individualist and structural approaches. One could also say that individualistic and structural approaches are compatible because they actually address different questions (Ayala & Vasilyeva 2015).

In this paper we combine elements of the two approaches in yet another way, by focusing on structural thinking - how individuals think about structures. It is an individualist proposal because it intervenes on individual's psychology, although not on implicit biases. In fact, Alex Madva (2016) points out that the individualistic approach is often defined too narrowly, and it gets identified with implicit bias. This restrictive definition results in treating failures of interventions on implicit bias as a failure of the whole approach. Our proposal contains individualistic elements that do

---

[1] We are not saying that Gigerenzer and other researchers working on heuristics endorse a structural approach to social injustice. We use Gigerenzer only to illustrate the idea that preventing cognitive shortcuts might not be the best way to go.

not imply intervene on implicit bias; instead, it is based on mitigating harmful internalist thinking. Our proposal can also be classified as structural, in two senses. First, it takes acknowledging the crucial role of structural factors to be the driving force for fighting injustice. Second, it can be implemented with structural-level interventions on human psychology. Before introducing the details of our proposal, we briefly review empirical literature on internalist thinking and how it contributes to discriminatory behavior.

2. Internalist thinking and its consequences

A growing body of psychological literature has documented the tendency for internalist thinking, the tendency to look "within" categories for causes and explanations, attributing properties of category members to inherent/essential factors (Cimpian & Salomon, 2014; Gelman, 2003). For example, if a person learns about a correlation between being an African American and higher likelihood of incarceration in the US, engaging in internalist thinking would lead them to attribute the property (propensity to end up in jail) to something about the category itself rather than the socioeconomic context. Just like an internalist thinker might expect that a tiger would naturally grow to be a ferocious predator with a similar set of predispositions no matter in what context it is brought up, they might expect that African Americans would achieve similar life outcomes even if they were placed in an entirely different social environment. Besides explanation, causal attribution, and intuitions about property

stability, internalist thinking can manifest itself in categorization, generalization, prediction, normative expectations and many other cognitive functions.

Psychological literature identifies several psychological phenomena that we place under the umbrella term of internalist thinking. Although they have important theoretical differences, for the purposes of our discussion we focus on commonalities, and on shared psychological and behavioral consequences. One example of internalist thinking is the inherence heuristic, defined as the tendency to explain observed patterns in terms of the inherent properties of the objects that instantiate the patterns (Cimpian & Salomon, 2014). For example, if girls wear pink, people might infer that it must be due to something inherent about pink (it is delicate) and/or girls (they are attracted to delicate colors), rather than considering a broader range of external, historical factors. This heuristic has been linked to a number of harmful psychological consequences, such as promoting support for the status quo (Hussak and Cimpian, 2015).

The inherence heuristic can be viewed as a precursor to the second phenomenon that we identify with internalist thinking, psychological essentialism, defined as the tendency to represent some natural and social categories as having underlying, deep, true natures (essences) which cause observable characteristics and ultimately make members of those categories what they are (Gelman, 2003; Medin & Ortony, 1989; Meyer et al., 2013; Hirschfeld, 1996; Prentice & Miller, 2007; Rhodes & Gelman, 2008; Rothbart & Taylor, 1992). It has been documented in children and adults, in reasoning about both natural and social kinds. Essentialized categories are represented as having unchanging and innate properties, and sharp and immutable boundaries, which can be

problematic when reasoning about both natural categories (e.g. interfering with understanding the mechanisms of natural selection; Shtulman and Schulz, 2008) and social categories (e.g., promoting unwarranted generalization of negative properties, strong normative expectations, prejudice, and contributing to the underrepresentation of women in some academic fields, such as mathematics, physics, engineering, and philosophy ; Allport, 1954; Haslam, Rothschild, & Ernst, 2000, 2002; Keller, 2005; Leslie, 2017; Leslie, Cimpian, Meyer, & Freeland, 2015).

Unfortunately, preventing development of essentialist beliefs appears to be a challenge; children readily acquire them from a rich variety of subtle cues, such as generic language (Cimpian & Markman, 2011; Rhodes, Tworek, & Leslie, 2012), e.g. statements attributing a property to a category in general, of the form "boys like X" and "girls do Y", which have been characterized as tagging categories as essence-based and generalization-supporting (Leslie, 2017). Children exposed to generic statements about an ability (even if the message is positive, such as "girls are good at X!") may show drops in motivation, interest, and performance (Cimpian, 2010; Cimpian, Mu, & Erickson, 2012). The essentialist stance appears to be quite resilient to evidence and revision. Non-generic quantified statements ("all", "most", "some") and ambiguous statements ("*they* are good at X") are often remembered and interpreted as generics by children and in some cases by adults (Hollander, Gelman & Star, 2002; Leslie & Gelman, 2012; Meyer & Baldwin, 2013; Tardif, Gelman, Fu, & Zhu, 2012). Even when adults do not openly endorse some essentialist beliefs (e.g., that girls "essentially" prefer dolls

over trucks), they may "default" to essentialist expectations when cognitively taxed (Eidson & Coley, 2014).

All this research paints a rather pessimistic picture for social change. Leslie (2014) is vocally skeptical about the possibility of breaking out of the loop of transmitting essentialist beliefs from generation to generation via subtle linguistic cues, where essentialist beliefs in turn shape social reality and make the essentialist beliefs true (in the sense that they reflect descriptive generalizations). Breaking out of this loop would seem to require going against deeply ingrained psychological tendencies to seek generalizable regularities and communicate about them efficiently. After all, we live in a world where boys *do* play with trucks, women are underrepresented in high-level academic positions in science, technology and philosophy, and African Americans are over-represented in prisons. It would be a heavy burden to put on people to ask them to put their capacity to learn about the environment on standby, to ignore salient statistical regularities, and to refrain from relying on their otherwise efficient cognitive tools that effectively support generalization and prediction in so many domains, such as associative learning and category-based reasoning.

Fortunately, there may be an alternative that doesn't require that people "un-see" stable correlations between categories and properties they observed in the world, or somehow discount these correlations as not meaningful. Instead, it would call for a different *explanation* of these connections, an explanation that would open the door for seeing these properties as not inevitable, but instead as depending on social *structures*.

3. Structural thinking

When philosophers, sociologists, psychologists or economists talk about the important role of socioeconomic structures shaping life outcomes for individuals, they engage in structural thinking. Its hallmark is locating an object of explanation within a larger structure and identifying *structural constraints* that act on components of the structure to shape the distribution of outcomes for each component (Haslanger, 2016). What we propose is that structural thinking is a broadly accessible cognitive skill, not restricted exclusively to trained professionals, and that promoting structural thinking in people could be an effective tool to mitigate the kind of injustice that concerns us here.

When applied to reasoning about social categories, structural thinking can offer an alternative explanation for why the observed correlations exist. For example, take properties such as "underrepresented in philosophy," "often quit their jobs after having children" – these can be reliably associated with the category "women" and taken to be true (in the sense of accurately describing the current state of the world). But, crucially, they may be represented not as caused by an internal, core property of "women," but as obtaining in virtue of women's position in the current socioeconomic structure (an "office," in Shapiro's terms (1997), or a "node," in Haslanger's terms (2016)).We refer to such properties of a category as *structural* properties, because they obtain in virtue of structural constraints acting on the "node" within a structure occupied by the category.

Looking at structural connections as a different type of non-accidental relationship between a property and a category brings to light an inherent ambiguity of generic statements and formal explanations (which explain a property by appealing to

category membership, e.g. "Why does Pazu walk over my computer keyboard while I work? - Because she is a cat"). Consider the following generics and formal explanations: "Women have trouble obtaining tenure in STEM fields"; "African American men end up in prison", "Ly didn't get tenure because she is a woman", "Willliam ended up in prison because he is African American". On most accounts, these statements are taken to express something about inherent properties of the respective categories (e.g., Cimpian & Markman, 2011; Leslie, 2014; Prasada & Dillingham, 2006, 2009); in contrast, we argue that it is possible to induce a non-essentialist, structural interpretation. Whereas internalist accounts represent features as characterizing the category per se, a structural account can represent features associated with a "node" or location within a structure.

Consider an example of a generic statement discussed by Haslanger (2011) and Leslie (2014): "women are submissive." This statement could be true due to external sociological factors (e.g., if the society systematically discourages women from being assertive). But, according to Leslie, this generic statement would nevertheless be interpreted as communicating that "*there is something in the nature of women* that makes them submissive" (p. 217), reinforcing essentialist beliefs about women. Leslie argues that generics are by default interpreted as expressing such "generalizations that hold because of common, inherent features of the members of the kind" (p. 217). The only alternative that Leslie considers is interpreting generics as describing statistical connections that do not reflect a principled relationship between a property and a category (Prasada & Dillingham, 2006, 2009), along the lines of "pigeons sit on statues,"

"police officers eat donuts," and "barns are red," on the basis of "specific worldly knowledge." In contrast, we focus on people's capacity to interpret generics and formal explanation structurally, by construing features of category members as products of structural constraints rather than inherent aspects of the kind or effects of causally active categorical essence.

Recent empirical evidence shows that with appropriate cueing, both children and adults can construe property-category connections structurally. Vasilyeva, Gopnik, and Lombrozo (2018) taught participants novel property-category connections ("girls play Yellow-ball", "boys play Green-ball"). Vasilyeva et al. varied the framing of the property-category connection: half of the participants were given structural cues (girls and boys were described as occupying structural positions with different affordances), and the other half of the participants were not given such cues. All participants (3-6-year-old children and adults) read a cover story about a school where each student received one of two toys (Yellow-ball or Green-ball) to play during recess; the toy was determined by whether a pebble the student threw fell into a yellow bucket or a green bucket. In one version of the story (designed to prime a structural construal), in the girl classroom the yellow bucket was much larger than the green bucket, with the reverse in the boys' classroom, imposing stable different structural constraints on the probability of each outcome for the girls and the boys in the story. In the other version of the story, the buckets were of the same size. Regardless of the story version, all participants received the same strong statistical evidence showing that across several days, most girls reliably played Yellow-ball, while most boys played Green-ball. Participants then

completed several tasks designed to measure how they construed the newly-learned

connection between the category (girl) and the property (playing Yellow-ball). First,

they were asked to explain the connection ("Why do girls play Yellow-ball?"); their

explanations were coded as structural—that is, appealing to differential accessibility of

the games for girls and boys (e.g., "Because for the girls, it is easier to get the pebble into

the yellow bucket")--or internalist–that is, appealing to category members' liking,

wanting, preferring one of the toys (e.g., "Because they like the color yellow")--or

miscellaneous. While in the absence of structural framing, most participants produced

internalist responses, when structural cues were given, the internalist responses

dropped with age as structural responses increased, so that by 5-6 years of age, the early

propensity to generate mostly internalist explanations switched to generating mostly

structural explanations in the structural condition. Strikingly, even the youngest group

(3-4 year-olds) was able to generate some structural explanations (about 20% of the

children in this age range did). Second, when participants intuition about property

stability were probed (in a scenario where a girl transfers to the boys' classroom, and

participants are asked if she would still play the Yellow-ball, even in the new context,

indicating low property mutability, or switch to playing the Green-ball, reflecting high

property mutability), for all age groups the structural framing of a property elicited

stronger expectations that the property can change (i.e. both children and adults were

significantly more likely to say that a girl would switch to playing a different game if

she is no longer confined to the "girl classroom", our equivalent of a structural node

with its fixed affordances). In the absence of structural priming participants expected

the category members to retain the property across contexts. These findings demonstrated that structural construal of category properties can be successfully induced across development (with stronger effects for adults, but – strikingly – with detectable effects with children as young as three), and that structural construal has tangible psychological consequences for how people explain properties of category members, and for the expectations they form about stability of those properties across possible contexts.

Another important finding from Vasilyeva, Gopnik, & Lombrozo (2018) was that there are tasks where the structural construal produces behavior indistinguishable from non-structural, internalist construal – but for very different reasons. When participants in the study were asked to evaluate a formal explanation explaining a property by appealing to the category membership – e.g. that Suzy plays Yellow-ball a lot at her school "because she is a girl" – participants were equally likely to endorse the explanation whether they committed to a structural or internalist construal of the property connection (as indicated by additional measures). This suggests that while both experimental groups agreed with the explanations, they meant different things by it – either that Suzy plays Yellow-ball a lot because she is a girl, and for girls Yellow-ball is more accessible whether or not they like it (the structural construal), or that Suzy plays Yellow-ball a lot because she's a girl, and girls are inherently drawn to yellow color / this particular type of toy (the internalist construal). This finding suggests that formal explanations (and, as ongoing work suggests, generic statements) come in two "flavors", internalist and structural, which opens up an intriguing possibility that

people with radically different views can seemingly agree while in practice talking past each other (as a bigoted person and a feminist can agree that "Mary didn't get tenure in math because she's a woman" – explaining the connection between women and difficulty of getting tenure in math either in terms of lack of innate talent, or in terms of structural barriers). On a more optimistic note, existence of the two "flavors" of formal explanations and generics suggests that despite all the negative consequences of generic speech documented in the psychological literature, we may not have to exorcise them from our speech – but we need to make sure we induce a structural interpretation when they are used. When someone states that "African Americans are over-represented in prisons", they need to emphasize the role of structural factors, unless they want to leave the door open for an internalist interpretation attributing the high odds of ending up in prison to inherent, deep characteristics of what it means to be an African American.

One potential benefit of structural explanations is their resilience, or the capacity to accommodate new statistical evidence about social categories without abandoning the explanation. In contrast, simply negating a property attributed to a social category may produce positive but short-lived effects, as new statistical information will likely serve as counterevidence to the negation. For example, if a person hears an assertion that previously established connection between a category ("women") and a property ("being submissive") is false - "that's not true, women are *not* submissive"[2] – when a

---

[2] Note that in light of flexible truth conditions of generics, simply adding a quantifier ("some women are submissive", "not all women are submissive") or specifying property prevalence numerically ("50% of women are submissive") may not be enough to make the generic "women are submissive" budge in its perceived truth value (Cimpian, Brandone, & Gelman, 2010). People are perfectly happy to endorse generics "lions have manes" and "ticks carry Lyme disease" while acknowledging that not all lions have manes, and only a small proportion of

new instance of a submissive behavior on the part of a woman is encountered, it will weigh in as falsifying the assertion. Similarly, when internalist *explanations* are negated without offering a structural alternative, they may relapse in light of new statistical evidence. For example, imagine a girl who learns that there are few women professors in mathematics departments, and asks a caretaker if that's because women are bad at it. The caretaker confidently denies that explanation asserting that actually, women are great at math. The girl may accept this claim, but when new evidence starts piling up (she learns that many women drop out after the first math course in college; few women receive prestigious awards in mathematics; the women she knows express little interest in math; etc.), it can function as contradictory evidence and an indication to revise her beliefs back to the same or a related internalist explanation (maybe after all women are not that good at math, and/or inherently not interested in it). In contrast, if the caretaker offered a structural explanation of the available evidence, attributing the dearth of women professors in mathematics to structural barriers women face in this discipline, then new observations may not be seen as questioning this explanation but compatible with it. At this point these claims about potential resilience of structural explanation are speculative, but one of the authors has relevant empirical studies in the works.

When applied to reasoning about causal relations, structural thinking can shift attention from triggering causes to structuring causes which make it so that the relationship between the triggering cause and the effect exists in the first place (Dretske,

---

ticks carry Lyme disease (Leslie, 2017; Tessler & Goodman, 2016). We are grateful to the editors for suggesting this clarification.

1988; Haslanger, 2016). There is nothing inherently wrong with being interested in triggering causes, but in social contexts where active structuring causes do shape problematic relationships between triggering causes and outcomes, ignoring structuring causes can be misguided. One example of focusing on a triggering cause is a 2013 popular article by Emily Yoffe telling college women to stop getting drunk since this behavior leads to sexual assault. In a response opinion piece, Louise Antony (2013) pointed out that that the relationship between women's drinking (the triggering cause) and sexual assault (the outcome) exists only in a particular type of society – in other words, she identified the social structure as the structuring cause that makes it so that the relationship "drinking → assault" exists in the first place, and suggested we should focus and intervene on structuring causes instead. The relationship between drinking and sexual assault is an example of an unstable relationship, defined as a relationship that varies in strength across background circumstances (Lewis, 1989; Woodward, 2006, 2010). Vasilyeva, Blanchard, & Lombrozo (2018) recently showed that when adults realize that a causal relationship is unstable, they are less likely to endorse causal and explanatory generalizations about them (of the form "X causes Y"), and less likely to act on the triggering cause to influence the outcome, compared to stable causal relationships. Pointing out structural causes shaping lower-level relationships between triggering causes and outcomes can make the instability of the latter salient, with tangible psychological consequences for reasoning about structurally-bound agents (i.e. most agents within a social structure). For example, having a child harms women's careers (Budig & England, 2001; Kahn, Garcia-Manglano, & Bianchi, 2014). Focusing on

the triggering cause (having a baby) puts the responsibility on women, with the corresponding low-level solution: advise women not to have children (at least until their careers are secured). Applying structural thinking and identifying structuring causes that shape the distribution of options for agents occupying different positions within the structure is an important step towards not blaming victims making constrained choices with constrained consequences within those social structures.

4. Structural thinking and epistemic injustice

How is structural thinking relevant for tackling epistemic justice? We see it as mitigating problematic processing of information about social groups that is commonly associated with epistemic injustice, in particular, stereotyping and prejudice. Essentialist beliefs, that we presented above as one element of internalist thinking, predict stereotype endorsement (Bastian & Haslam 2005) and are associated with prejudice (Allport 1954; Haslam, Rothschild, Ernst 2002). When attributes of social groups are seen as fixed and derived from an internal core, people are likely to rely on stereotypes to navigate the social reality. As long as structural thinking takes us away from the sort of cognitive processes that lead to problematic judgements of other people, developing structural thinking has something to offer to the fight against epistemic injustice.

Specifically, structural thinking can help reduce hermeneutical injustice by addressing conceptual gaps that align with larger patterns of discrimination. Structural thinking might be the missing hermeneutical tool for understanding statistical patterns in non-discriminatory ways (e.g. offering a non-internalist account of why certain social

groups are overrepresented or underrepresented in specific domains, without assuming inevitability and normativity of the existing patterns). It thus puts thinkers in a better position to reason about the social categories they and others belong to. This has direct consequences for testimonial and distributive injustice. When one reason a person A discounts a person B as a potential knower is because A engages in an internalist interpretation of B's "failures" in other domains (e.g. failing to secure certain jobs or levels of income) by attributing them to an inherent lack of capacity or knowledge on part of B or a social category B belongs to – structural thinking may neutralize this chain of inferences by offering non-discriminatory ways of interpreting associations in the world (e.g. between B's social category and low wages), with positive downstream consequences for evaluation of B's testimony (B is perceived as nevertheless credible and trustworthy). Even if this is not sufficient to eliminate perceived credibility deficit entirely, by training A to consider non-internalist explanations of the (apparent) lack of credibility, structural thinking can provide A with the necessary conceptual tools, the hermeneutical resources required to appreciate B's capacity to contribute knowledge. In cases when the listener lacks detailed knowledge about the speaker, structural thinking may prevent or reduce an internalist interpretation of the initial disfluency, or lack of immediate comprehension of what the speaker is saying, or more generally, the perceived trustworthiness of the speaker, by offering an alternative, non-internalist interpretation of these observations (e.g, A might think "perhaps, I don't understand what B is saying (or B does not appear credible to me) not because B is inherently un-comprehensible, but because the mismatch in the social positions each of us occupies

makes communication between us more challenging"). In this sense, thinking structurally might pave the way to a better epistemic sensibility, to the capacity to tune up to unfamiliar hermeneutical systems and unfamiliar knowers, a "Kaleidoscopic social sensibility" (Medina 2013).

How does the structural thinking proposal compare to previously articulated interventions against epistemic injustice in terms of feasibility? For example, when Fricker talks about the virtue of epistemic justice, she writes that "Its possession requires the hearer to reliably neutralize prejudice in her judgments of credibility" (Fricker 2007, 92). As others have pointed out, developing and exercising the virtue of epistemic justice seems like a difficult enterprise. Fricker proposes a sort of epistemic leap of faith. That is, when we judge a knower in a faulty way, and this is due to our prejudices presenting the knower as a deficient source, then we need to resist what appears as the evidence. When our epistemic prejudices-mediated detectors are telling us to be cautious, or outright skeptical, we need to trust. The virtue of epistemic justice seems to be telling us to ignore the (apparent) evidence we are getting, and trusting that we will know how to make a correct judgement in the absence of it. At a first glance, it sounds like the opposite of epistemic responsibility, and yet we know this is something we would do out of both epistemic and moral responsibility: because we want to get the reality right in spite of the smoke curtain of prejudices distorting it. Even if not impossible, resisting the prejudice-mediated evidence sounds like a difficult thing to do.

The virtue of epistemic justice also requires that we have the capacity to notice

when we have made a mistaken judgement about a knower. The idea behind epistemic justice is that once we notice it, then we compensate for the mistake. As pointed out in Sherman (2015), there are problems with committing to this capacity. From the first-person perspective, I might have the impression that I have become good at detecting faulty judgements that derive from my prejudices. However, how can I be sure about the reliability of that evidence? As Sherman explains, some errors, such as the gambler's fallacy, can be fixed with a critical, reflective attitude. While it is easy for us to commit that fallacy, after we reflect we might realize of the mistake, and we can try to correct for it. He doubts, however, that we can identify and correct for the mistakes that have to do with our prejudiced judgements. What would count as evidence of the mistake if our perception is prejudiced in the first place?

In comparison, we see structural thinking as not requiring that sort of psychological heroism apparently required by the virtue of epistemic justice. Instead of going against the apparent evidence or calling for detecting mistakes that by definition escape our radar (as they can occur at the level of implicit bias), structural thinking requires that we consider an alternative explanation for the association between the category and the properties, such as perceived lack of credibility. We can speculate that structural thinking may be easier to implement than the virtue of epistemic justice, for two reasons. It appears that the virtue of epistemic justice requires that an agent attributes the perceived lack of credibility either to some unknown cause, or to the agent's own cognitive and/or moral flaws. But we know from the psychological literature that both of these strategies may be an uphill battle. First, people have trouble

reasoning about causal models with unspecified inputs, where "nothing" appears to cause "something". For example, when people are told that a medical screening test produces false positives in 5% of the times (but are not told why or how), they fail to apply this knowledge when interpreting test results (Eddy, 1982; Gigerenzer & Hoffrage, 1995). Asking people to acknowledge the possibility that their "detectors of un-trustworthy and non-credible speakers" produce false positives may be similarly ineffective. Interestingly, giving people a concrete variable as the cause of false positives (e.g. saying that they can be caused by a benign cyst) significantly improves performance (Krynski & Tenenbaum, 2007); once the concrete variable is plugged into the mental causal model as the cause, it is easier for people to take its effect into account, since the effect does not appear to arise "out of nothing". But, in the case of epistemic justice virtue approach, an equivalent move hits another barrier: people tend to resist attributing negative causal influences to self (see reviews of evidence for self-serving bias, Blaine & Crocker, 1995; Miller & Ross, 1975; and actor-observer asymmetry, Jones & Nisbett, 1971). In contrast, the structural thinking approach provides the agent with a tangible *external* cause of the perceived lack of credibility, without threatening the agent's self. This, taken together with the fact that structural thinking does not require actively discarding available statistical evidence, combined with the evidence of structural thinking in children as young as three (Vasilyeva, Gopnik, & Lombrozo, 2018), gives us reasons to believe that structural thinking is a psychologically plausible approach, and it may turn out to be less cognitively demanding than employing the virtue of epistemic justice.

5. When structural thinking is (un)likely to help

Of course, structural thinking is not a magic cure-all solution, and it is important to determine when it is and is not likely to work. One possibility is that while it succeeds in blocking overgeneralization across social contexts, it may be less resistant to overgeneralization across category members[3]. Prior research on non-structural generics documented an asymmetry between generic truth conditions and implied prevalence: in one study (Cimpian, Brandone, & Gelman, 2010; Experiment 1), participants learned that a certain percentage of novel category members had a property (e.g., "30% of lorches have purple feathers") and indicated whether the corresponding generic ("lorches have purple feathers") was true or false. On average, participants required 69% or more category members to have the property in order to endorse the generic as true. However, when a different group of participants was given the same generic statements and asked to estimate the percentage of category members having the property, their estimates averaged to 96%. This showed that generics can be endorsed based on little evidence, but have strong implications – potentially resulting in overgeneralization across individual category members. Can structural thinking help mitigate this effect? The short answer is we do not know yet (but one of the authors is currently conducting studies to evaluate the effects of structural construal on generic truth conditions and implied prevalence; Vasilyeva & Lombrozo, in prep.). Depending on the psychological mechanisms giving rise to the asymmetry between the generic truth and implied prevalence judgments, structural framing of generics can prove more

---

[3] We are grateful to the editors of this volume for encouraging us to develop this question.

or less impactful. Let's consider three possible psychological mechanisms. First, people could inflate prevalence estimates because they assume that a novel generic conveys a principled feature of a category (a la "dogs have tails"; Prasada & Dillingham, 2006, 2009), rather than a striking or a minority characteristic ("sharks attack swimmers"; "lions have manes"; Leslie, 2007, 2008); they might make this assumption if they believe that generics about principled features are the most common or prototypical (Cimpian, Brandone, & Gelman, 2010). If so, then making a structural generic alternative more salient to people can help mitigate inflation of prevalence estimates in two ways: by helping them recognize that the proportion of "principled generics" is lower than they thought, and by highlighting that structural generics may hold true of mutable – and thus potentially more variable across individuals – properties (Vasilyeva, Gopnik, & Lombrozo, 2018). On the other hand, if the asymmetry stems from pragmatic assumptions (the speaker's choice not to quantify the generic with "some" may be taken to signal that all category members rather than a subset have the property, based on the Gricean maxim of quantity; Cimpian, Brandone, & Gelman, 2010; Declerck, 1991; Grice, 1975) – then the structural framing may have little impact. As a third option, the asymmetry may stem from the previously documented discrepancy between outputs of extensional reasoning (targeting the extension of the concept, the set of items it picks out in the world) and intensional reasoning (about the abstract kind and its features), and the specific challenges of translating between the two, in either direction (Cimpian, Brandone, & Gelman, 2010; Jönsson & Hampton, 2006; Tversky & Kahneman, 1983). If so, structural thinking might complicate the matters by requiring people to manage

uncertainty over the abstract kind in question: for example, when considering the generic "women are submissive", are we talking about "women" qua an essentialized kind, or "women" qua a structural node (Haslanger, 2016)? If the two kinds disagree in prevalence estimates of submissiveness, the mere act of acknowledging the uncertainty may produce more conservative prevalence estimates, mitigating overgeneralization across individuals.

Apart from the issue of erroneous (inflated) estimates of feature prevalence in a category, there is the question of moral permissibility of relying on *statistically accurate* estimates about social groups to guide decisions about their members. On the one hand, ignoring base rates, or prior probabilities of a person having a feature, is considered a violation of principles of rationality, which undermines prediction accuracy (Bar-Hillel, 1980; Tversky & Kahneman, 1974). For example, in a medical context, in order to maximize accuracy of diagnosis for a patient with a cough, a symptom compatible with both the common cold and lung cancer, doctors should take into account that there are a lot more people with the common cold than there are people with lung cancer; ignoring this information will result in dramatic deviations from accurate diagnosis, under-diagnosing common colds and mis-diagnosing lung cancer.

Likewise, social stereotypes have been conceptualized as offering useful base rate information that maximizes prediction accuracy (Krosnik, Li, & Lehman, 1990; Locksley, Borgida, Brekke, & Hepburn, 1980; Rasinski, Crocker, & Hastie, 1985). However, the use of social stereotypes for prediction is not morally neutral: it can be argued to go against principles of justice and fairness, violating rights of individuals

(Cao & Banaji, 2016; Cao, Kleiman-Weiner, & Banaji, 2017; Dworkin, 2000; Rawls, 2001). This idea is reflected in explicit prohibitions against using group membership and corresponding base rates in the US courts to evaluate defendants' guilt, and to adjust health insurance premiums in Europe (Cao, Kleiman-Weiner, & Banaji, 2017; Koehler, 1992; European Court of Justice, 2011). In recent work, Cao, Kleiman-Weiner and Banaji (2018) show that people can do both – behave like rational Bayesians, using base rate information to maximize accuracy of their predictions, and condemn the use of base rate information on moral grounds. For example, the same group of participants who predicted that, e.g., a man who performed a surgery is more likely to be a doctor compared to a woman who performed a surgery, also judged a person making the same type of a Bayesian judgment as unfair and unjust[4].

We can speculate that structural thinking might encourage people to take into account the considerations of justice and fairness when they make decisions about how to act towards members of different social groups, by drawing their attention to the

---

[4] Cao, Kleiman-Weiner and Banaji (2018) raise a possibility that participants are not being hypocritical (essentially accusing a third person of bigotry for the very same judgment they made themselves), but instead they engage in another rational Bayesian judgment, based on participants' estimates of base rates of sexists in our society. Consistent with our proposal about the "two flavors" of generics, internalist and structural, "both an unabashed sexist and a feminist statistician can state that a man who performed surgery is more likely to be a doctor than is a woman who performed surgery, albeit for different reasons" (Cao, Kleiman-Weiner, & Banaji, 2018, p. 10). Thus, participants had to estimate whether the third party was a sexist or a feminist. If they believe that there are more sexists than feminists, it would be rational to assume the speaker was a sexist, and criticize them. Ultimately, the sexist is blameworthy for arriving to an accurate conclusion in a wrong way. This is another illustration of the psychological importance attributed to the different causal mechanisms, even if they lead to the same outcome. (Additionally, this suggests yet another potential effect of engaging in structural thinking: a listener who recognizes that speakers may use generics and formal explanations to convey structural meanings will be less likely to jump to the conclusion that anyone who uses generic language about gender is sexist, and will avoid false alarms classifying structural thinkers as bigots).

causes of the observed statistical patterns across social categories, helping them recognize role of the stable patterns of disadvantage. Extensive psychological evidence suggests that people care deeply about the causal mechanisms, or causal models producing the observed outcomes (Griffiths & Tenenbaum, 2005; Tenenbaum, Griffiths, & Kemp, 2006); when they realize that several possible causal models could have produced the observation, even young children take steps to collect additional evidence to figure out which model was responsible (Schulz & Bonawitz, 2007). When people generalize from limited observations, different explanations of the same observation produce radically different generalization behaviors (Lombrozo & Gwynne, 2014; Vasilyeva & Coley, 2013). Thus, we can expect structural vs. internalist explanations of an observed statistical correlation between a category and a property to influence how people reason and act based on this correlation.

To illustrate this idea, imagine you "hire" two teenage kids living in your neighborhood to help you deliver free lunches to some residents. It turns out that both fail to deliver about 30% of the lunches, but for different reasons: one is lazy, unmotivated and distracted, so she loses lunch boxes or forgets to deliver them; the other one is regularly bullied by neighborhood bullies, who destroy her lunchboxes when they encounter her on her delivery route. Assuming the same expected outcome in the future (30% loss), would you be equally likely to hire each of these kids again? Our intuition suggests that the preference might go to the second kid, with a more structural explanation of the negative outcome, rather than the first kid, with a more internalist explanation of the same outcome. Likewise, we can speculate that a CEO

choosing between two candidates with the same average likelihood of underperforming may be sensitive to the underlying mechanism, making them more likely to hire the candidate whose failures arise due to structural factors (e.g. misses evening meetings due to lack of available childcare, does not pursue new projects due to lack of support from co-workers) than the candidate failing for intrinsic reasons (e.g. forgets to attend meetings due to lack of organization, does not pursue new projects due to lack of motivation)[5]. In sum, if the explanatory mechanism could make a difference in how much weight is given to moral considerations of justice and fairness, then structural thinking has the potential to make a difference in how members of some social groups are treated, even if the predictive statistics do not speak in their favor[6].

Were structural thinking to prove beneficial for alleviating epistemic (and other typs of) injustice, it could be promoted at several levels. At the first-person level, it can be implemented as a way of self-improvement. It can also be promoted by encouraging others to engage in it. Just like Vasilyeva, Gopnik, and Lombrozo (2018) were able to

---

[5] Non-structural explanations that merely cite external factors may be less efficient in mitigating prejudice. For example, Belief in Social Determinism (BSD), a lay theory that "a person's essential features…are shaped *permanently* and *profoundly* by social factors" (Rangel & Keller, 2011, p. 1, our emphasis), utilizes external factors to tell an "inside story" of how individuals come to be "permanently and profoundly" shaped by social factors (Plaut & Markus, 2005): e.g. how a woman came to be submissive. Critically, the consequences of engaging in BSD are not that different from internalist thinking: it is associated with dispositional bias, expectations of stability across situations and time, negative stereotyping, prejudice, discrimination, and hierarchy-enhancing ideologies (e.g., nationalism). Applying it to our examples, the BSD explanation would appeal to the social factors that made one of the candidate lazy, irresponsible and unmotivated. While alleviating the potential blame, we doubt this explanation would do much in boosting the candidate's odds of being hired.

[6] Of course, there are limits on how powerful an explanation can be in guiding action. When a choice puts moral and social responsibility in conflict with productivity and commercial success – e.g., if the choice is between a candidate likely to underperform for structural reasons and a candidate very unlikely to underperform at all – the conflict may be resolved differently, depending on the "mission" and priorities of the hiring organization. In academia, selection of graduate students or research assistants may raise similar conflicts in decision-makers.

promote structural thinking in their participants, teachers could promote structural thinking in the classroom, by engineering the kind of inquiry contexts children confront. This suggests yet another level of implementation, in the form of guidelines about how best to talk to children, especially when generic statements are involved.

6. Concluding notes

The proposal we have outlined combines the individualist with the structural elements in a novel way: it targets how individuals think about structures. By acknowledging that both the individuals' psychology and the structural factors contribute to producing injustice, it engages both of these elements in fighting it, in an empirically-informed way. Our proposal also calls for further empirical research to evaluate effectiveness and boundary conditions of structural thinking.

References

Anthony, L. (2013). Advising women against drinking also sends a dangerous method, *New York Times* October 24, 2013.

https://www.nytimes.com/roomfordebate/2013/10/23/young-women-drinking-and-rape/advising-women-against-drinking

Allport, G. W. (1954). *The Nature of Prejudice*. Cambridge, MA: Addison-Wesley.

Ayala, S. & Vasilyeva, N. (2015). Explaining Speech Injustice: Individualistic vs. Structural Explanation, in D.C. Noelle, R. Dale, A.S. Warlaumont, J. Yoshimi, T. Matlock, C.D. Jennings, and P.P. Maglio (eds.) *Proceedings of the 37th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica, 44,* 211–233.

Bastian, B. & Haslam, N.(2005). Psychological essentialism and stereotype endorsement. *Journal of Experimental Social Psychology*, *42*, 228–235

Blaine, B., & Crocker, J. (1993). Self-esteem and self-serving biases in reactions to positive and negative events: An integrative review. In *Self-esteem* (pp. 55-85). Springer, Boston, MA.

Budig, M.J., & England, P. (2001). The wage penalty for motherhood. *American Sociological Review, 66*, 204–225.

Cao, J. & Banaji, M. R. (2016). The base rate principle and the fairness principle in social judgment. *Proceedings of the National Academy of Science. USA* 113, 7475–7580.

Cao, J., Kleiman-Weiner, M., & Banaji, M. R. (2018). People Make the Same Bayesian

Judgment They Criticize in Others. *Psychological science*,

https://doi.org/10.1177/0956797618805750

Cimpian, A. (2010). The impact of generic language about ability on children's

achievement motivation. *Developmental Psychology*, *46(5),* 1333-1340.

http://dx.doi.org/10.1037/a0019665

Cimpian, A., Brandone, A. C., & Gelman, S. A. (2010). Generic statements require little

evidence for acceptance but have powerful implications. *Cognitive science*, *34*(8),

1452-1482.

Cimpian, A., & Markman, E. M. (2011). The generic/nongeneric distinction influences

how children interpret new information about social others. *Child Development,*

82, 471–492. http://dx.doi.org/10.1111/j.1467-8624.2010.01525.x

Cimpian, A. Mu, Y. & Erickson, LC. (2012). Who is good at this game? Linking an

activity to a social category undermines children's achievement. *Psychological*

Science, 23(5), 533-41. doi: 10.1177/0956797611429803.

Cimpian, A., & Salomon, E. (2014). The inherence heuristic: An intuitive means of

making sense of the world, and a potential precursor to psychological

essentialism. *Behavioral and Brain Sciences*, *37,* 461–480.

http://dx.doi.org/10.1017/S0140525X13002197

Davidson, L. and Kelly, D. (manuscript). Minding the Gap: Bias, Soft Structures, and the

Double Life of Social Norms.

Declerck, R. (1991). The origins of genericity. *Linguistics, 29*, 79–102.

Dotson, K. (2012). A cautionary tale: On limiting epistemic oppression, *Frontiers: A Journal of Women Studies, 33 (1)*:24-47

Dretske, F. (1988). *Explaining behavior: Reasons in a world of causes.* Cambridge, MA: MIT Press.

Dworkin, R. (2000). *Sovereign Virtue: The Theory and Practice of Equality*. Harvard Univ. Press, Cambridge.

Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 249 –267). Cam- bridge, England: Cambridge University Press.

Eidson, R. C., & Coley, J. D. (2014). Not so fast: Reassessing gender essentialism in young adults. *Journal of Cognition and Development*, *15(2)*, 382-392. http://dx.doi.org/10.1080/15248372.2013.763810

European Court of Justice (2011). *Test-Achats v. Council of Ministers.*

Fricker, M. (2007). *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought.* New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780195154061.001.0001

Gigerenzer, G. (2010) Moral Satisficing: Rethinking Moral Behavior as Bounded Rationality. *Topics in Cognitive Science, 2,* 528-554.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102,* 684 –704.

Grice, H. P. (1975). *Logic and conversation*. In P. Cole & J. L. Morgan (Eds.), Syntax and semantics, Vol. 3 (pp. 41–58). New York: Academic Press.

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive psychology*, *51*(4), 334-384.

Haslam, N., Rothschild, L., & Ernst, D. (2002). Are essentialist beliefs associated with prejudice? British Journal of Social Psychology, 41, 87–100.

Haslam, N., Rothschild, L., & Ernst, D. (2000). Essentialist beliefs about social categories. *British Journal of Social Psychology, 39* (Pt 1), 113-27.

---- (2002). Are essentialist beliefs associated with prejudice? *British Journal of Social Psychology*, 41(Pt 1), 87-100.

Haslanger, S. (2017). Injustice within Systems of Coordination and Cognition. Comment on Madva for Brains Blog, The Brains Blog, March 6, 2017.

Haslanger, S. (2016). What is a (social) structural explanation?. *Philosophical Studies*, *173*(1), 113-130.

Hirschfeld, L. A. (1996). *Learning, development, and conceptual change. Race in the making: Cognition, culture, and the child's construction of human kinds.* Cambridge, MA, US: The MIT Press.

Hollander, M., Gelman S., Star J. (2002). Children's interpretation of generic noun phrases. *Developmental Psychology, 38,* 883–894

Hussak, L. J., & Cimpian, A. (2015). An early-emerging explanatory heuristic promotes

    support for the status quo. *Journal of Personality and Social Psychology, 109*, 739 –

    752. http://dx.doi.org/10.1037/pspa0000033

Jones, E. & Nisbett, R. (1971*). The actor and the observer: Divergent perceptions of the causes*

    of behavior. New York: General Learning Press.

Jönsson, M. L., & Hampton, J. A. (2006). The inverse conjunction fallacy. *Journal of*

    Memory and Language, 55, 317–334.

Kahn, J. R., García-Manglano, J., & Bianchi, S. M. (2014). The Motherhood Penalty at

    Midlife: Long-Term Effects of Children on Women's Careers. *Journal of marriage*

    and the family, *76(1),* 56-72.

Keller, J. (2005). In genes we trust: the biological component of psychological

    essentialism and its relationship to mechanisms of motivated social

    cognition. *Journal of personality and social psychology*, *88*(4), 686.

Koehler, J. (1992). Probabilities in the Courtroom: An Evaluation of the Objections and

    Policies, in Kagehiro, D. & Laufer, W. (Eds) *Handbook of Psychology and Law* (pp.

    167–184) Springer, New York.

Krosnick, J. A., Li, F. & Lehman, D. R. Conversational conventions, order of

    information acquisition, and the effect of base rates on individuating social

    information on social judgments. *J. Pers. Soc. Psychol.* **59**, 1140–1152 (1990).

Krynski, T. R., & Tenenbaum, J. B. (2007). The role of causality in judgment under

    uncertainty. *Journal of Experimental Psychology: General*, *136*(3), 430.

Leslie, S.-J. (2007). Generis and the structure of the mind. *Philosophical Perspectives, 21*(1), 375-405.

Leslie, S.-J. (2008). Generics: Cognition and acquisition. *The philosophical Review, 117*(1), 1-49.

Leslie, S.-J. (2014). Carving up the social world with generics. In T. Lombrozo, J. Knobe, & S. Nichols (Eds.), *Oxford studies in experimental philosophy* (pp. 208–232). New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780198718765.003.0009

Leslie, S.J. (2017). The original sin of cognition: Fear, prejudice and generalization. The *Journal of Philosophy*, *114*, 393–421.

Leslie, Cimpian, Meyer, & Freeland, .(2015). Expectations of brilliance underlie gender distributions across academic disciplines, *Science* 16 Jan 2015, 347 (6219),262-265

Leslie, S.-J., & Gelman, S. A. (2012). Quantified statements are recalled as generics: evidence from preschool children and adults. *Cognitive Psychology*, *64*(3), 186–214. doi:10.1016/j.cogpsych.2011.12.001

Locksley, A., Borgida, E., Brekke, N. & Hepburn, C. Sex stereotypes and   social judgment. *J. Pers. Soc. Psychol.* **39**, 821–831 (1980).

Lombrozo, T., & Gwynne, N. Z. (2014). Explanation and inference: mechanistic and functional explanations guide property generalization. *Frontiers in human neuroscience*, *8*, 700.

Madva Alex. 2016. A plea for anti-anti-individualism: How oversimple psychology misleads social policy. *Ergo* 3(27), 701-728.

Medin D.L., & Ortony A. (1989). Psychological essentialism. In: Vosniadou S, Ortony A, editors. *Similarity and analogical reasoning* (pp. 179–196). Cambridge University Press; New York, NY.

Medina, J. (2013). *The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and Resistant Imaginations*, Oxford: Oxford University Press.

Meyer M., Leslie SJ, Gelman SA, & Stilwell SM. (2013). Essentialist beliefs about bodily transplants in the United States and India. *Cognitive Science*, 34, 668–719.

Meyer, M. & Baldwin, .D. (2013). Pointing As a Socio-Pragmatic Cue to Particular vs.Generic Reference. *Language Learning and Development*, 9(3), 245-265.

Miller, D. T., & Ross, M. (1975). Self-serving biases in the attribution of causality: Fact or fiction?. *Psychological bulletin*, *82*(2), 213.

Plaut, V. C., & Markus, H. R. (2005). The "inside" story: A cultural-historical analysis of being smart and motivated, Americal style. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 457-488). New York, NY: Guilford Press.

Pohlhaus, G. Jr. (2012). Relational Knowing and Epistemic Injustice: Toward a Theory of Willful Hermeneutical Ignorance, *Hypatia: A Journal of Feminist Philosophy*, *27(4)*, 715–735.

Prasada, S., & Dillingham, E. M. (2006). Principled and statistical connections in common sense conception. Cognition, 99, 73–112. http://dx .doi.org/10.1016/j.cognition.2005.01.003

Prasada, S., & Dillingham, E. M. (2009). Representation of principled connections: A

> window onto the formal aspect of common sense conception. *Cognitive Science*,
>
> *33,* 401–448. http://dx.doi.org/10.1111/j.1551-6709.2009.01018.x

Prentice, D. A., & Miller, D. T. (2007). Psychological Essentialism of Human

> Categories. *Current Directions in Psychological Science*, *16*(4), 202–
>
> 206. https://doi.org/10.1111/j.1467-8721.2007.00504.x

Rangel, U., & Keller, J. (2011). Essentialism goes social: Belief in social determinism as a

> component of psychological essentialism. *Journal of Personality and Social*
>
> *Psychology, 100*(6), 1056-1079. doi:10.1037/a0022401

Rasinski, K. A., Crocker, J. & Hastie, R. (1985). Another look at sex stereotypes and

> social judgments: an analysis of the perceiver's use of subjective probabilities. *J.*
>
> *Pers. Soc. Psychol.* 49, 317–326.

Rawls, J. (2001). *Justice as Fairness: A Restatement*, Harvard Univ. Press, Cambridge.

Roberts, D. E. (2004). The Social and Moral Cost of Mass incarceration in African

> American Communities. *Faculty Scholarship*. Paper 583.

Rhodes, M., & Gelman, S. A. (2008). Categories influence predictions about individual

> consistency. *Child Development*, *79*(5), 1270-1287.

Rhodes, M., Leslie, S.J., & Tworek, C.M. (2012).Cultural transmission of social

> essentialism. *Proceedings of the National Academy of Sciences*, *109(34),* 13526-31. doi:
>
> 10.1073/pnas.1208951109.

Rothbart, M., & Taylor, M. (1992). Category labels and social reality: Do we view social categories as natural kinds? In G. Semin & K. Fiedler (Eds.), *Language, interaction and social cognition* (pp. 11–36). London: Sage.

Schulz, L. E., & Bonawitz, E. B. (2007). Serious fun: preschoolers engage in more exploratory play when evidence is confounded. *Developmental psychology*, *43*(4), 1045.

Shapiro, S. (1997). *Philosophy of mathematics: Structure and ontology.* Oxford: Oxford University Press.

Sherman, B. (2015). There's No (Testimonial) Justice: Why Pursuit of a Virtue is Not the Solution to Epistemic Injustice. *Social Epistemology* 30(3), 229-250. doi: 10.1080/02691728.2015.1031852

Shtulman A., & Schulz, L.(2008). The Relation Between Essentialist Beliefs and Evolutionary Reasoning. *Cognitive Science, 32,* 1049–1062.

Tardif, T., Gelman, S. A., Fu, X., & Zhu, L. (2012). Acquisition of generic noun phrases in Chinese: Learning about lions without an '-s'. *Journal of Child Language*, *39*(1), 130-161.

Tessler, M. H., & Goodman, N. D. (2016). A pragmatic theory of generic language. *arXiv preprint arXiv:1608.02926*.

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in cognitive sciences*, *10*(7), 309-318.

Tversky, A. & Kahneman, D. (1979). The framing of decisions and the psychology of choice. *Science 211(4481),* 453-458.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90(4),* 293–315.

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science 185,* 1124–1131.

Vasilyeva, N., Blanchard, T.J., & Lombrozo, T. (2018). Stable causal relations are better causal relations. Cognitive Science, 42(4), 1265-1296.

Vasilyeva, N., & Coley, J. D. (2013). Evaluating two mechanisms of flexible induction: selective memory retrieval and evidence explanation, in *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, eds M. Knauff, M. Pauen, N. Sebanz, and I. Wachsmuth (Austin, TX: Cognitive Science Society), 3654–3651.

Vasilyeva, N., Gopnik, A., & Lombrozo, T. (2018). The development of structural thinking about social categories. Developmental Psychology, 54(9), 1735-1744.

Vasilyeva, N., & Lombrozo, T. (in prep.) Generic truth conditions and implied prevalence: the effects of structural explanation.

Yoffe, E. (2013). College women: Stop getting drunk. *Slate*, October 15, 2013. https://slate.com/human-interest/2013/10/sexual-assault-and-drinking-teach-women-the-connection.html

Zheng, R. (2018). Bias, Structure, and Injustice: A Reply to Haslanger. *Feminist Philosophy Quarterly* 4, (1). Article 4. doi:10.5206/fpq/2018.1.4.